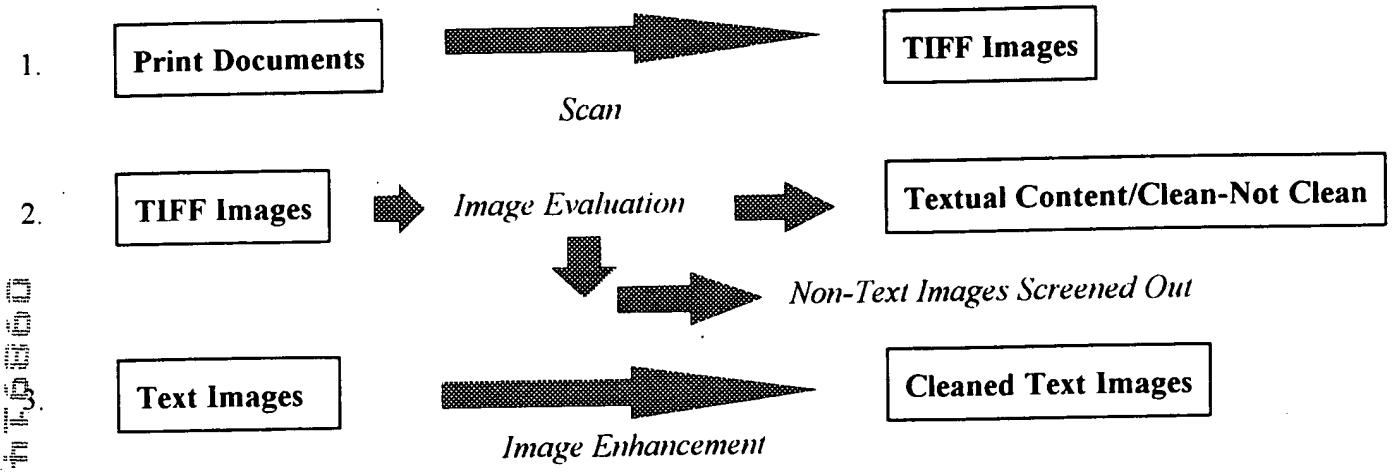
Figure 1

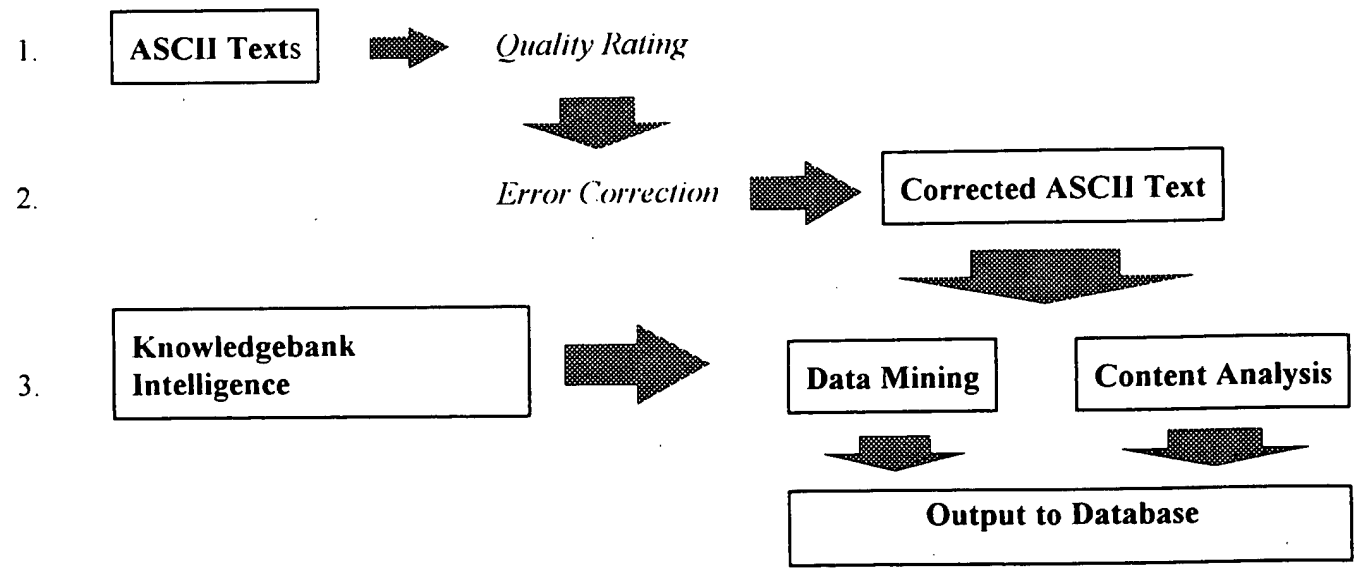
Image-to-Data Processing System:

Pre-OCR Processes



Optical Character Recognition: Image-to-Text Conversion

Post-OCR Processes: Text-to-Data Conversion



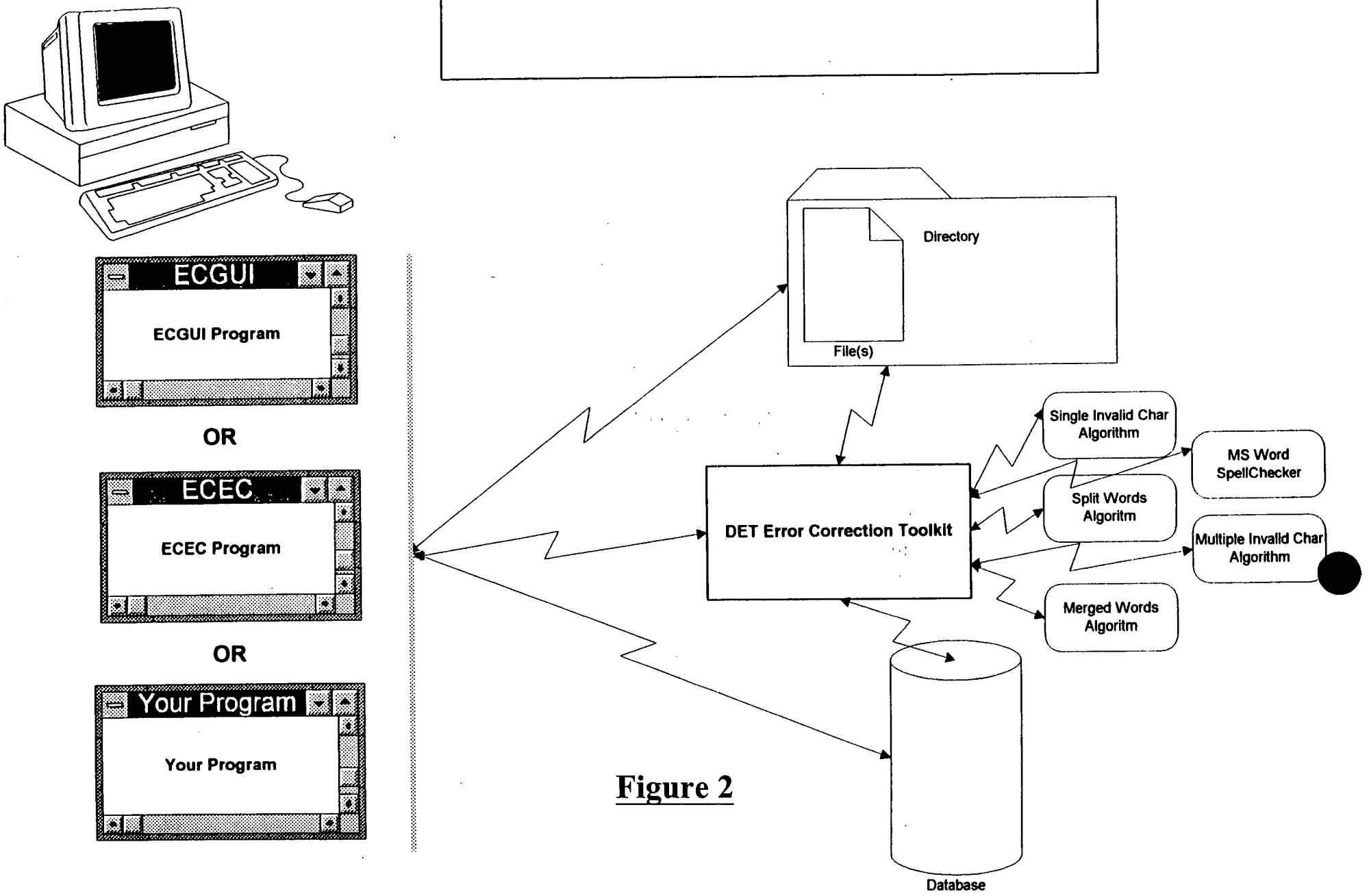

ECGUI
ECGUI Program
OR
ECEC
ECEC Program
OR
Your Program
Your Program
Directory
File(s)
DET Error Correction Toolkit
Single Invalid Char Algorithm
MS Word SpellChecker
Split Words Algoritm
Multiple Invalid Char Algorithm
Merged Words Algoritm
Database

Figure 2

Computer-Aided Error Correction

Figure 3

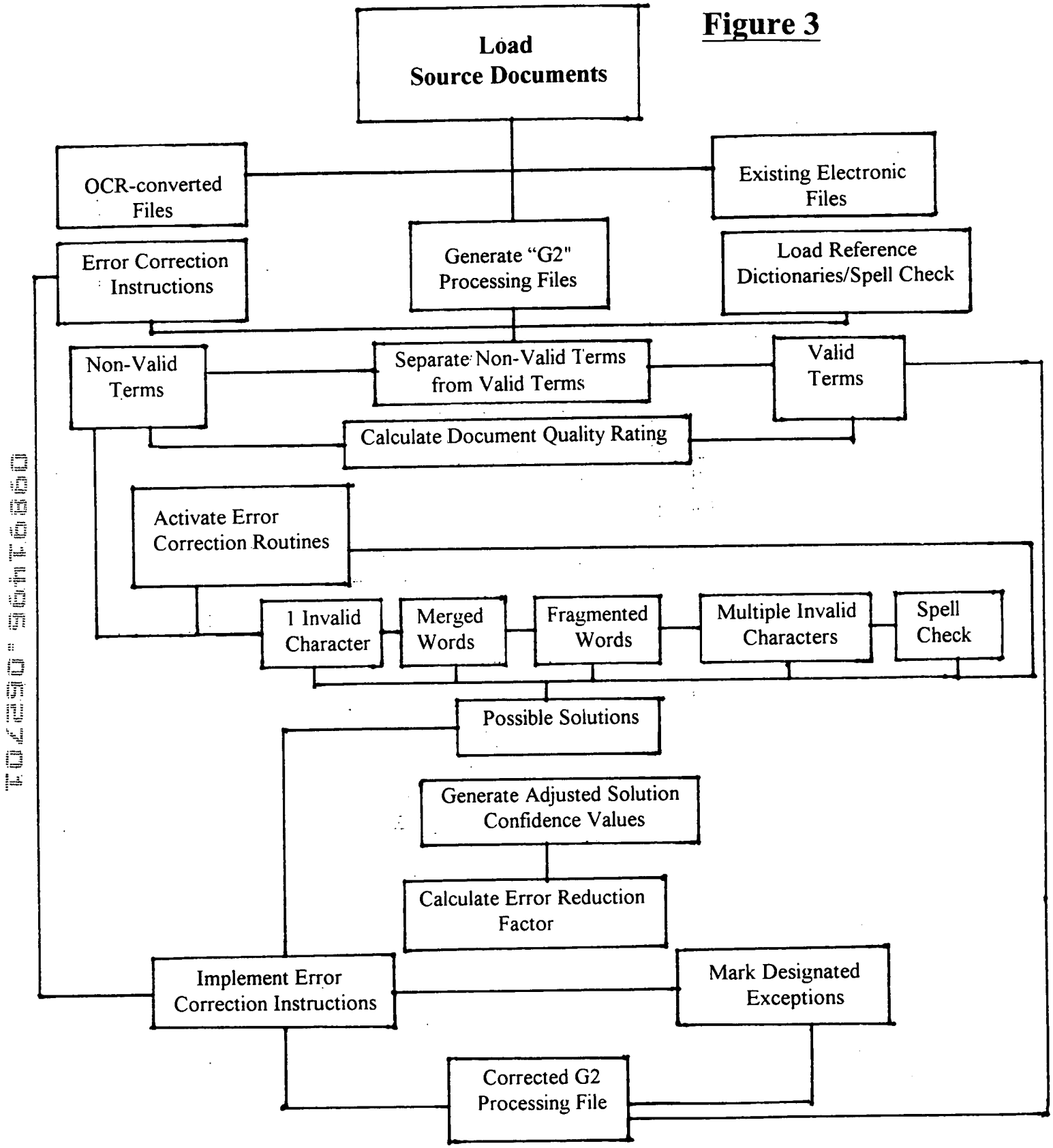


Figure 4

Diagram: Screening

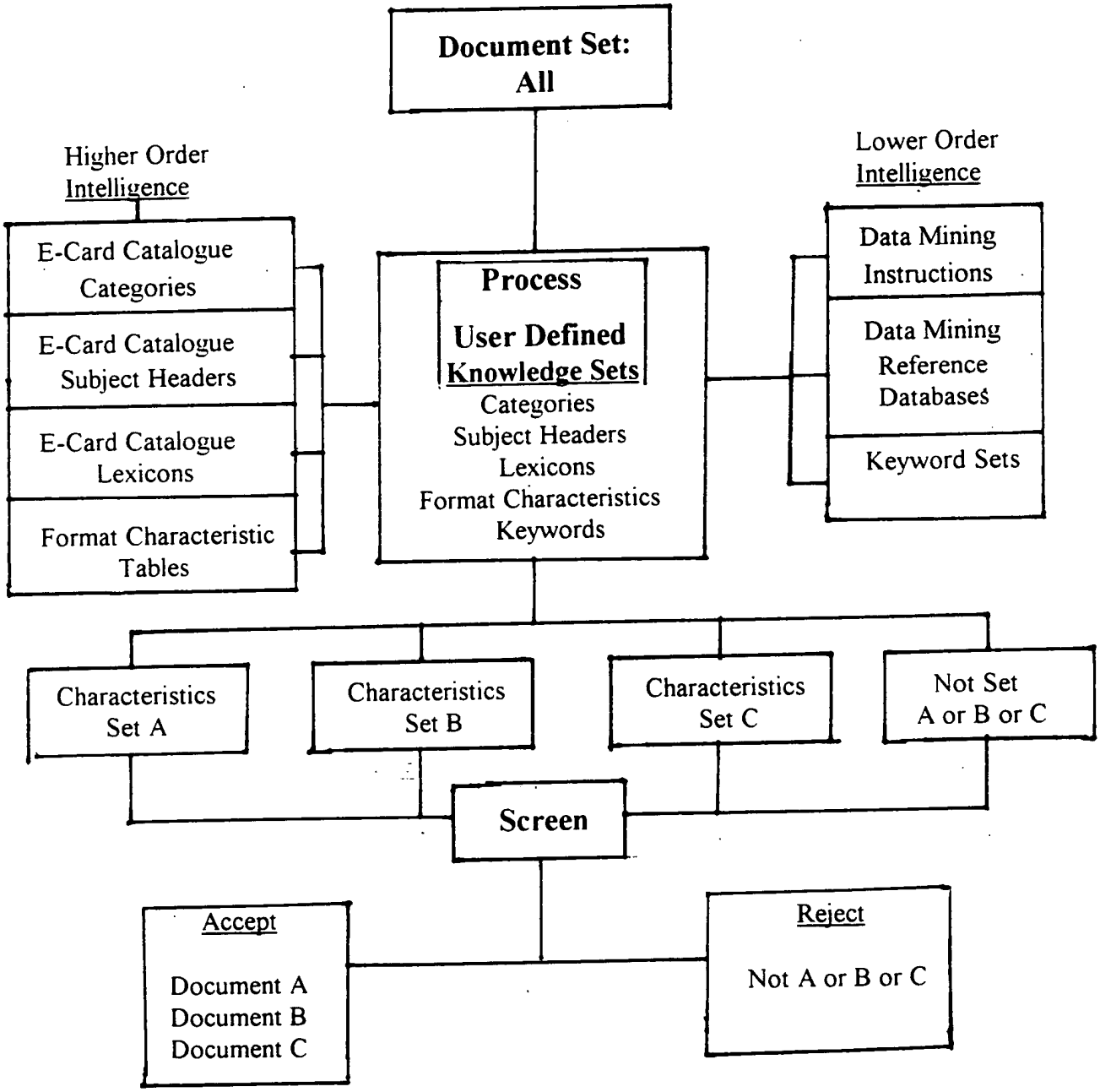


Figure 5

Diagram: Sorting

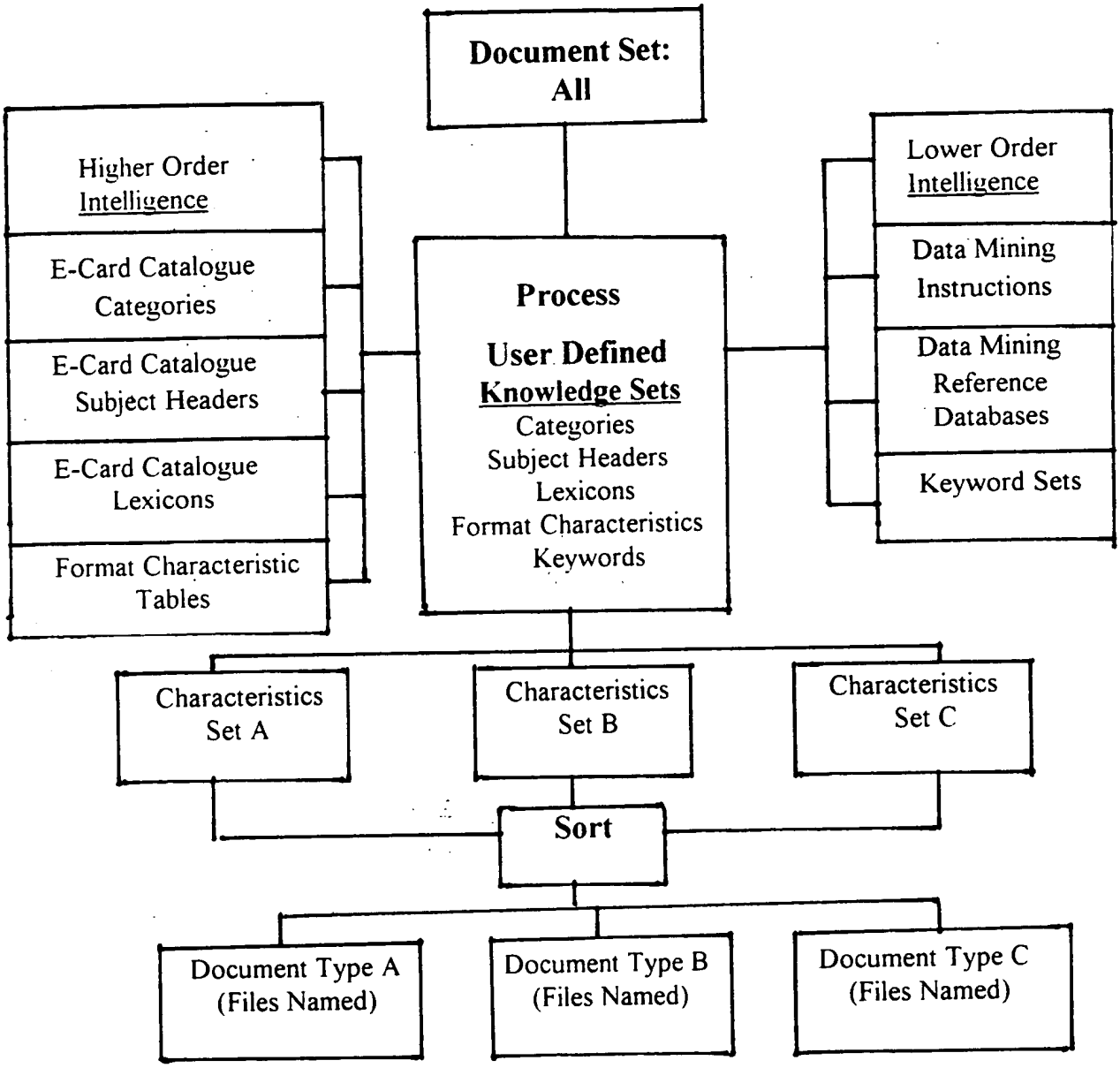


Figure 6

Diagram: Characterizing

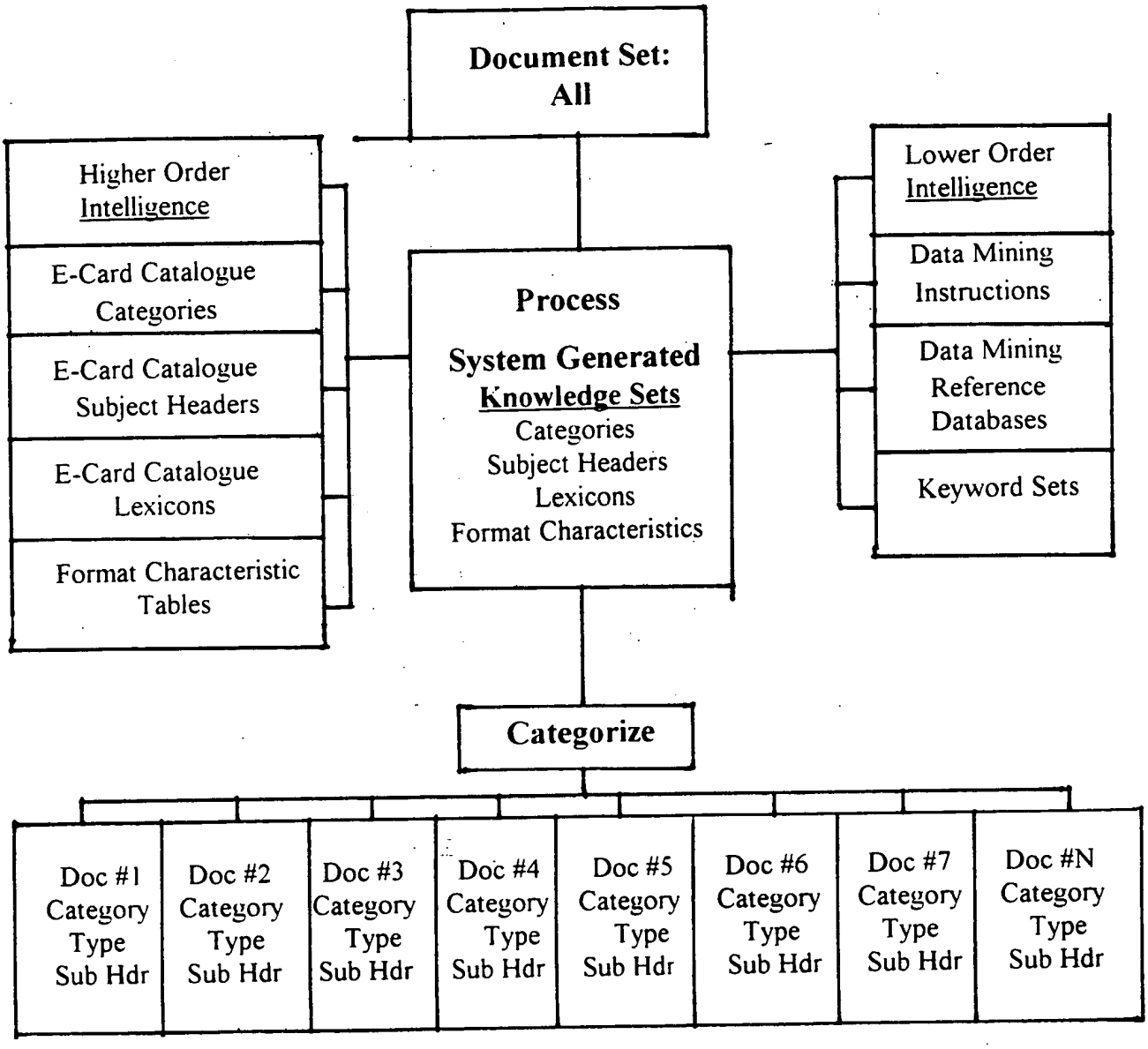
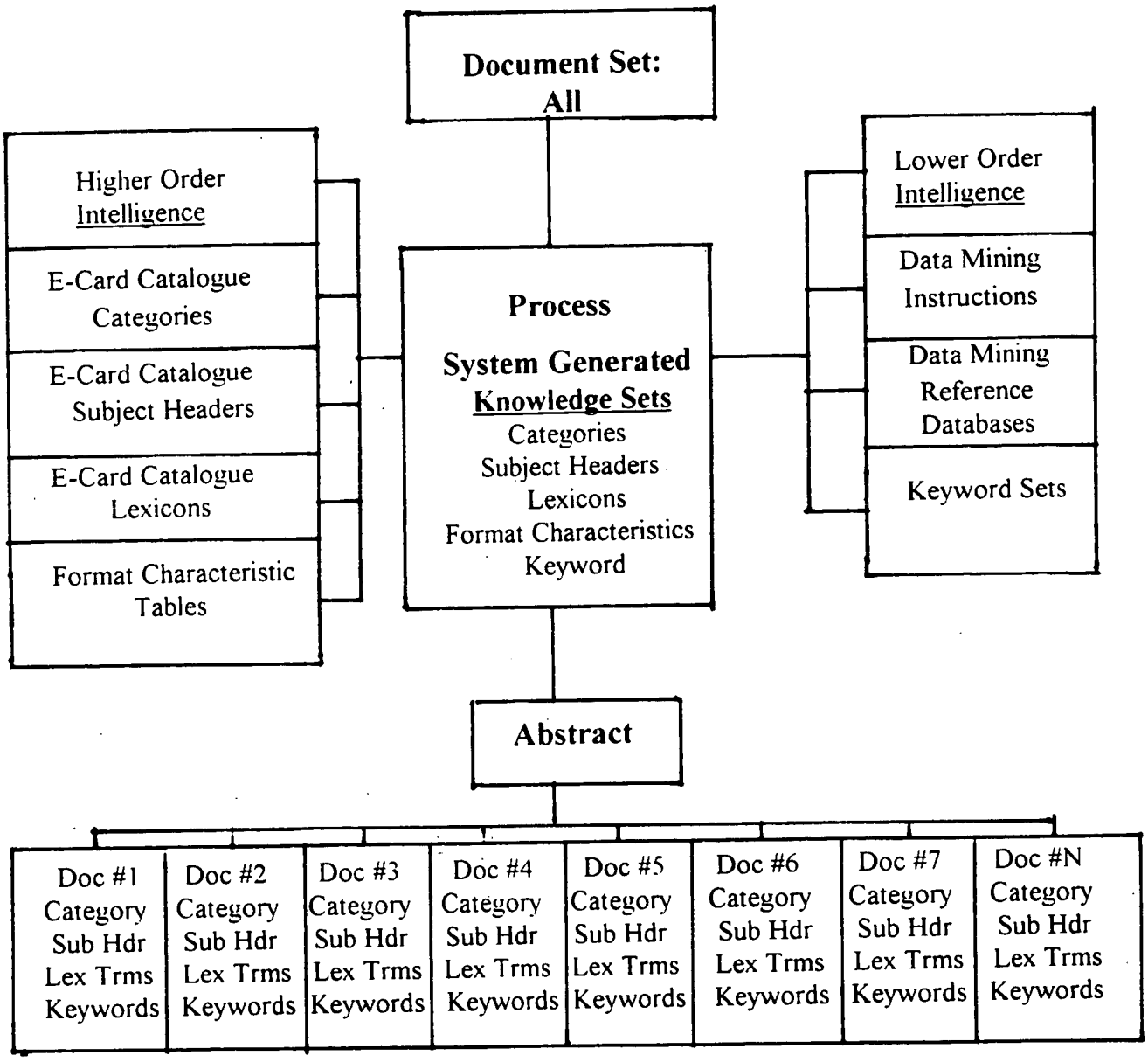


Figure 7

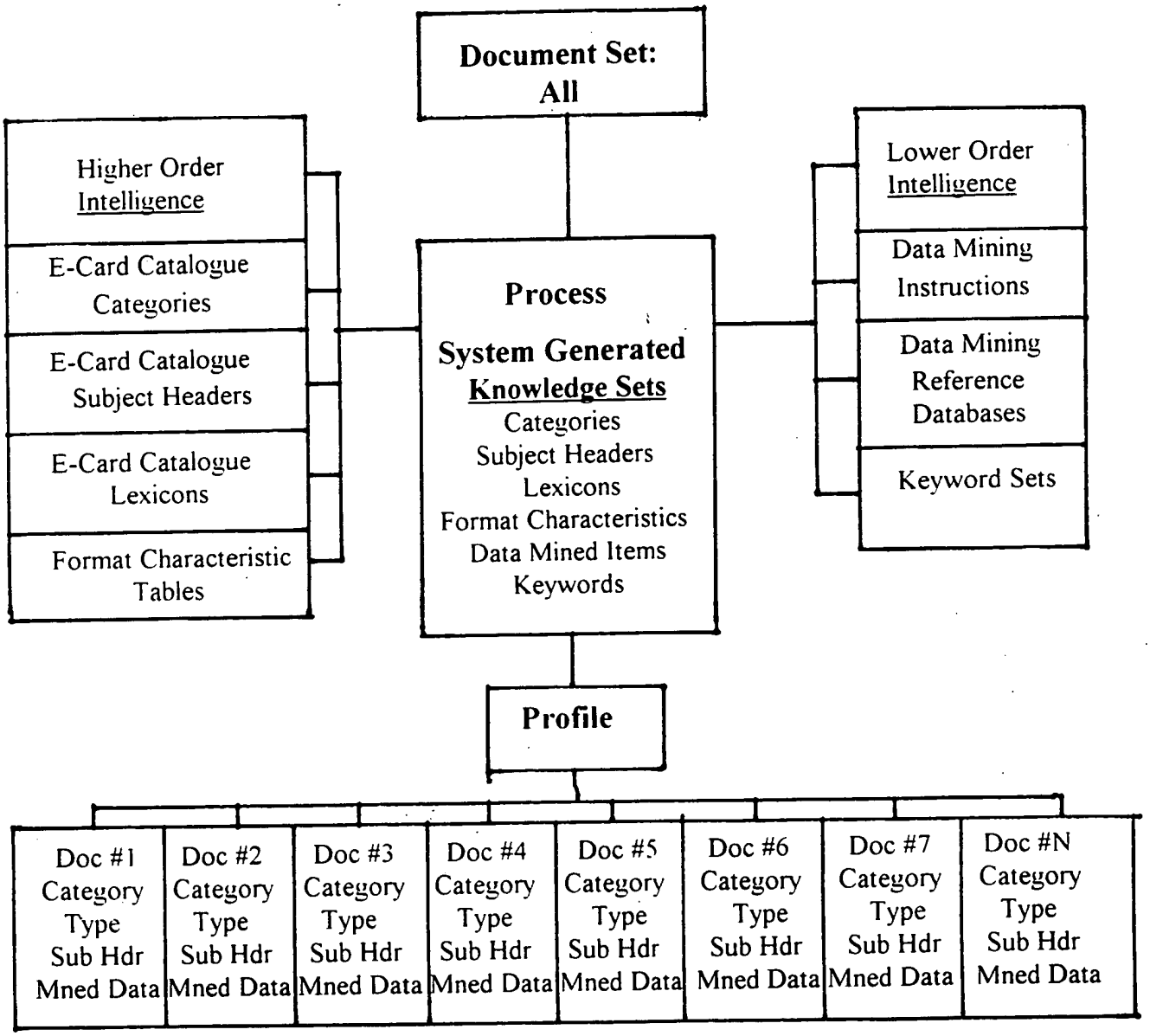
Diagram: Abstracting



EXPRESS MAIL NO. EV 339 150 405 US

Figure 8

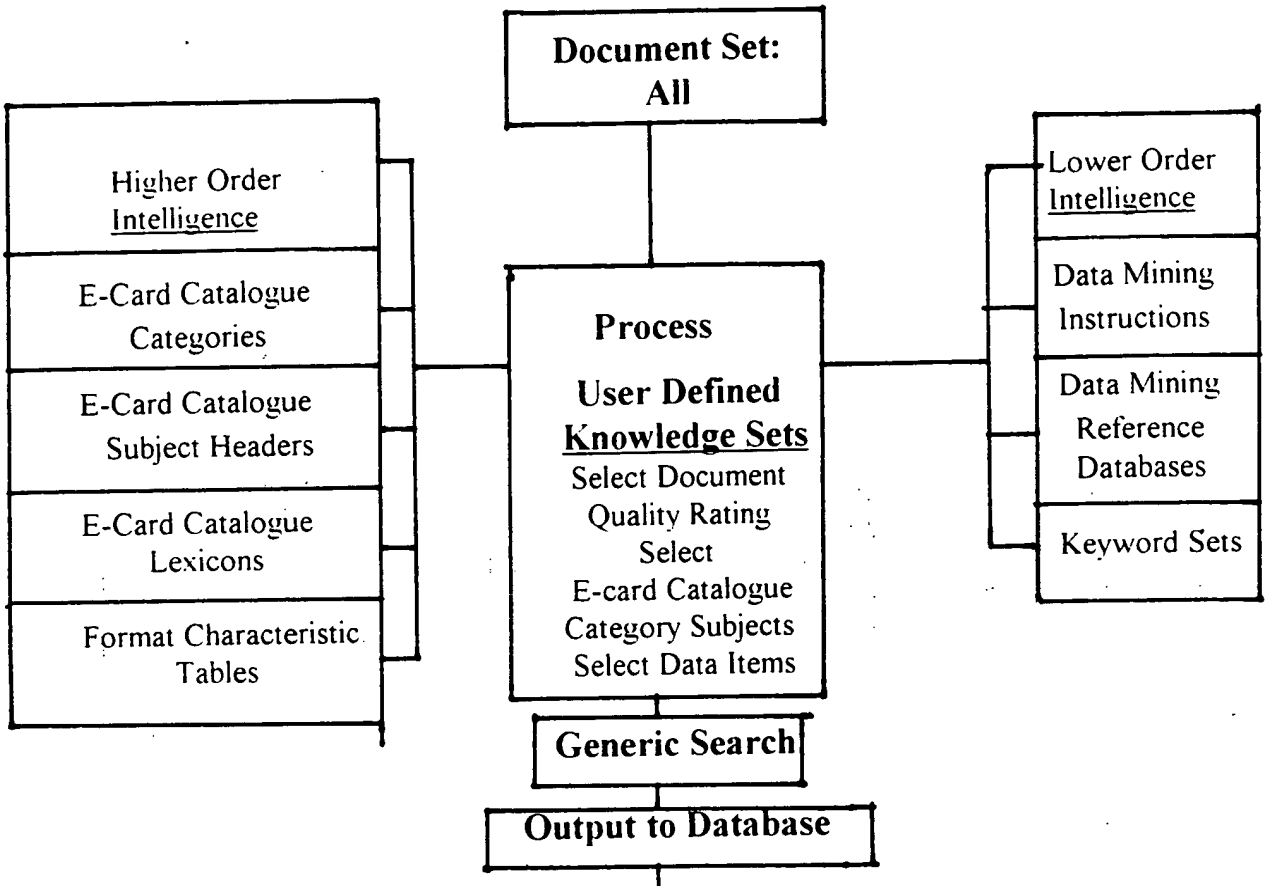
Diagram: Profiling



Express Mail Label No. EV 339 150 905 US

Figure 9

Diagram: Generic Searching

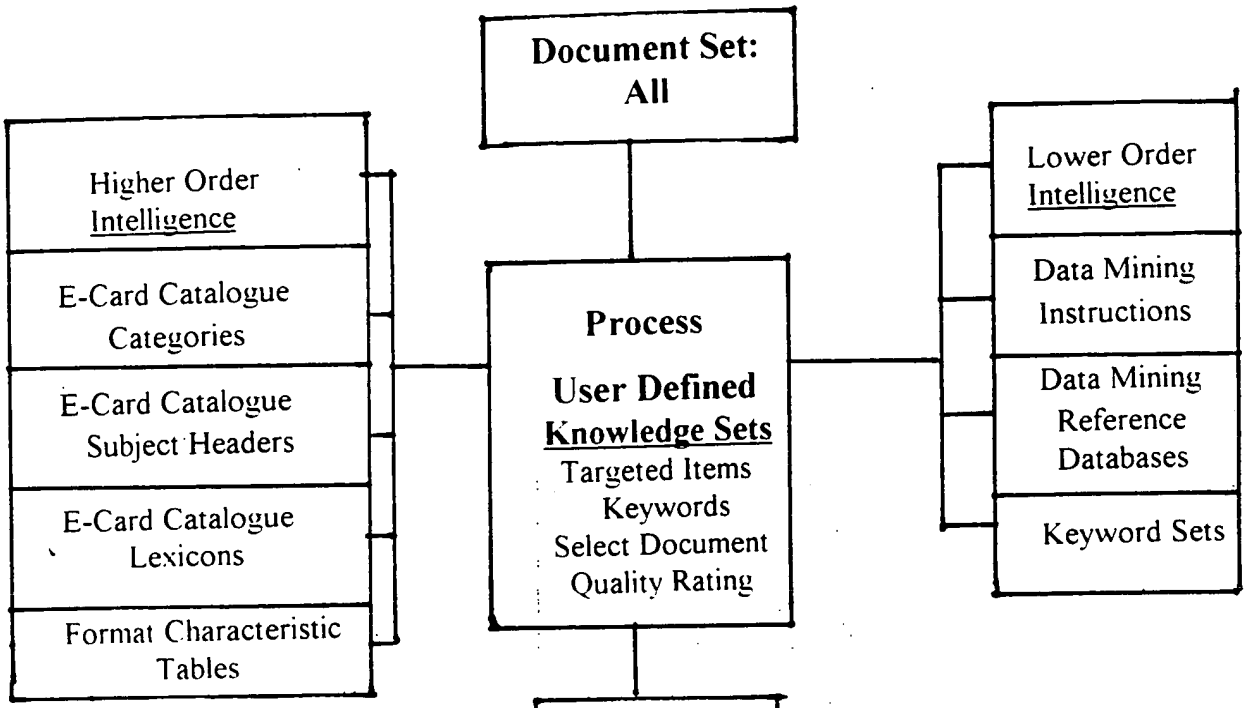


Names	Medical Terms	Doc #1* Body Parts	Professional Designations
Names	Medical Terms	Doc #2* Body Parts	Professional Designations
Names	Medical Terms	Doc #3* Body Parts	Professional Designations
Names	Medical Terms	Doc #4* Body Parts	Professional Designations
Names	Medical Terms	Doc #5* Body Parts	Professional Designations
Names	Medical Terms	Doc #6* Body Parts	Professional Designations
Names	Medical Terms	Doc #7* Body Parts	Professional Designations

* as an example

Figure 10

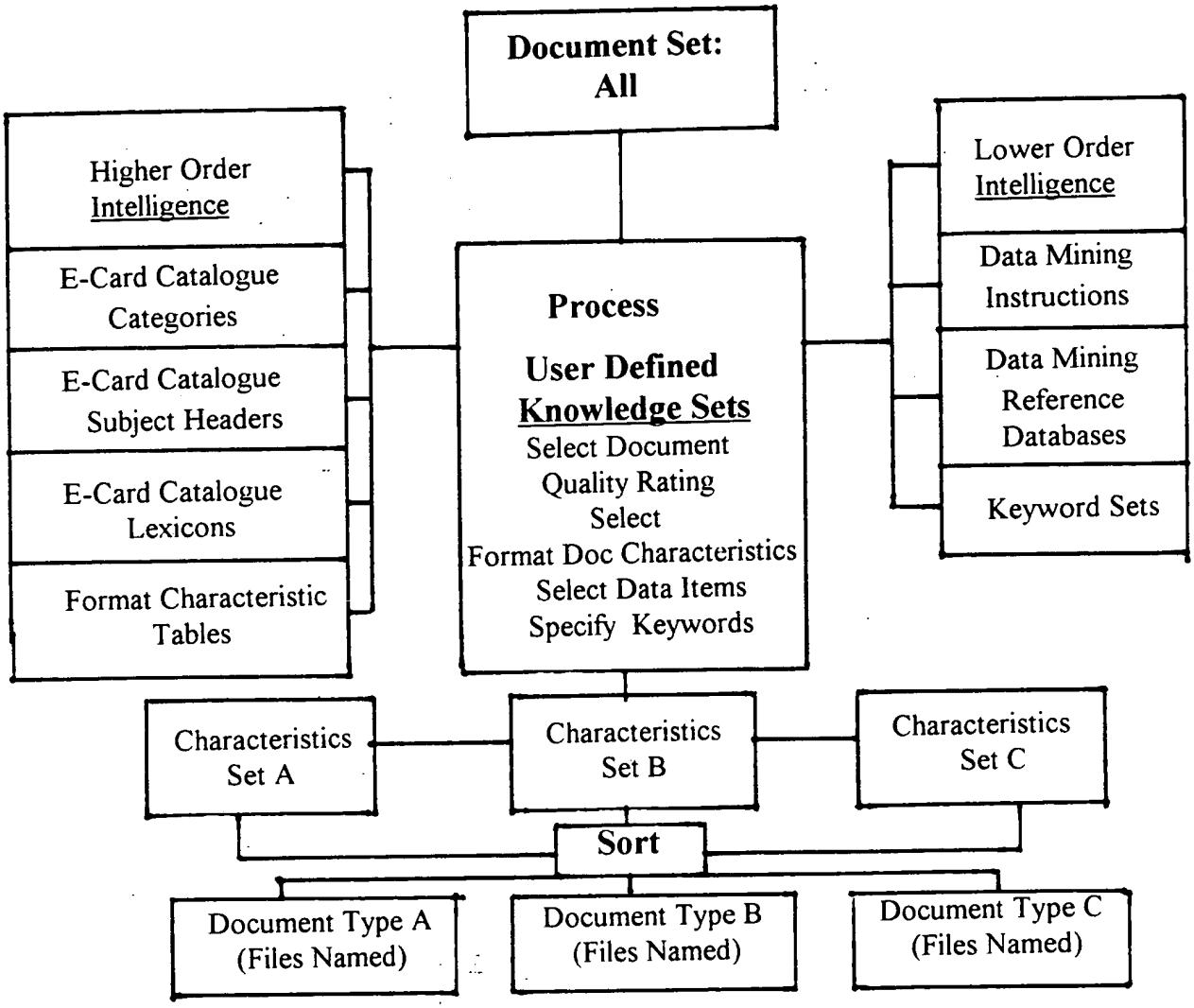
Diagram: Data Mining



Doc #1				
Target Item A	Target Item B	Target Item C	Target Item D	Target Item E
Doc #2				
Target Item A	Target Item B	Target Item C	Target Item D	Target Item E
Doc #3				
Target Item A	Target Item B	Target Item C	Target Item D	Target Item E
Doc #4				
Target Item A	Target Item B	Target Item C	Target Item D	Target Item E
Doc #5				
Target Item A	Target Item B	Target Item C	Target Item D	Target Item E
Doc #6				
Target Item A	Target Item B	Target Item C	Target Item D	Target Item E
Doc #7				
Target Item A	Target Item B	Target Item C	Target Item D	Target Item E

Figure 11

Diagram: Non Standard Forms Processing



Mine Data Appropriate to Document Type

Document Type A	Document Type B	Document Type C
(File Name)	(File Name)	(File Name)
Mined Data	Mined Data	Mined Data
(File Name)	(File Name)	(File Name)
Mined Data	Mined Data	Mined Data
(File Name)	(File Name)	(File Name)
Mined Data	Mined Data	Mined Data

Output to Database

Figure 12

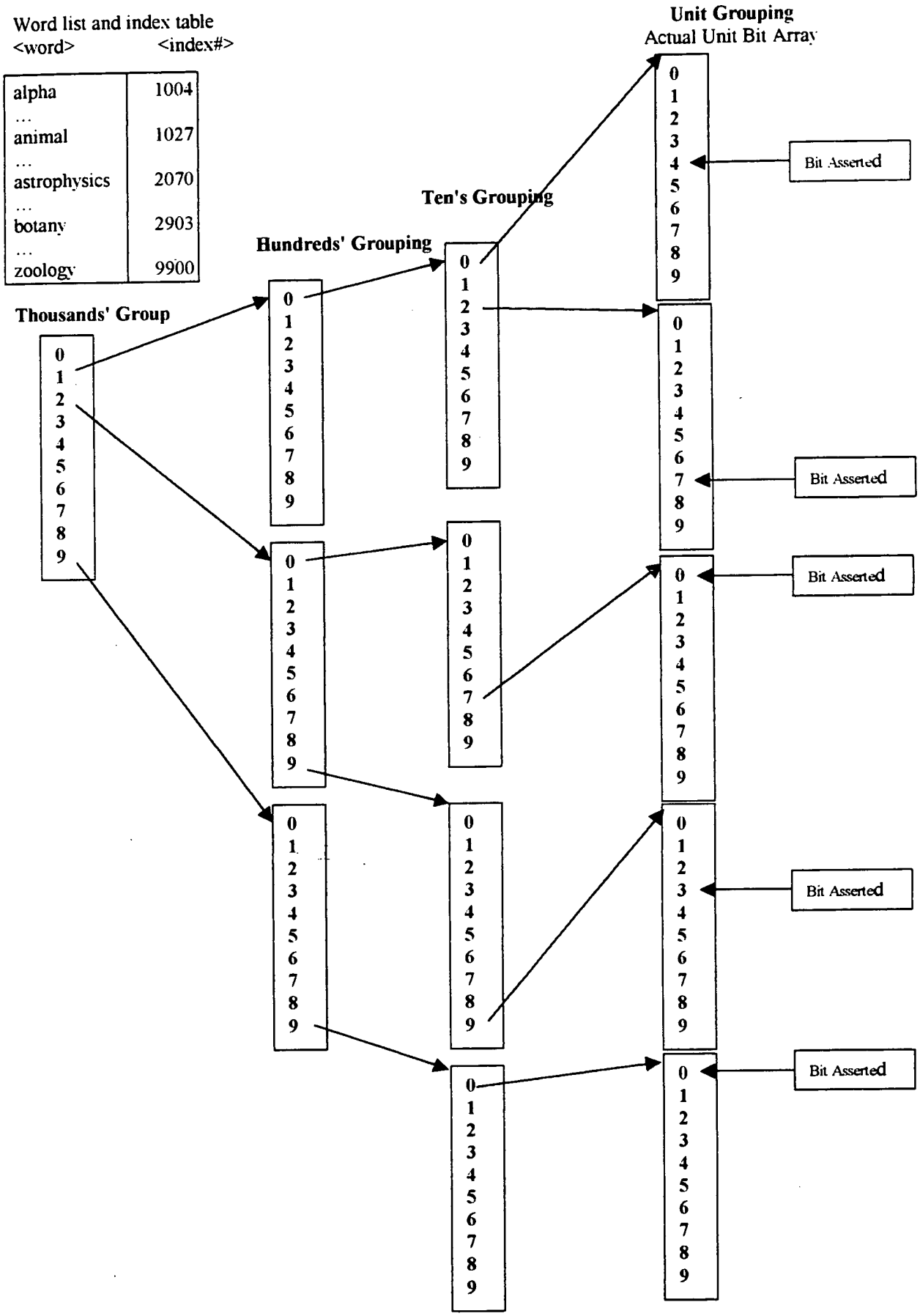

Word list and index table
<word>
<index#>
alpha 1004
animal 1027
astrophysics 2070
botany 2903
zoology 9900
Thousands' Group
Hundreds' Grouping
Ten's Grouping
Unit Grouping
Actual Unit Bit Array
Bit Asserted
Bit Asserted
Bit Asserted
Bit Asserted
Bit Asserted

Figure 13

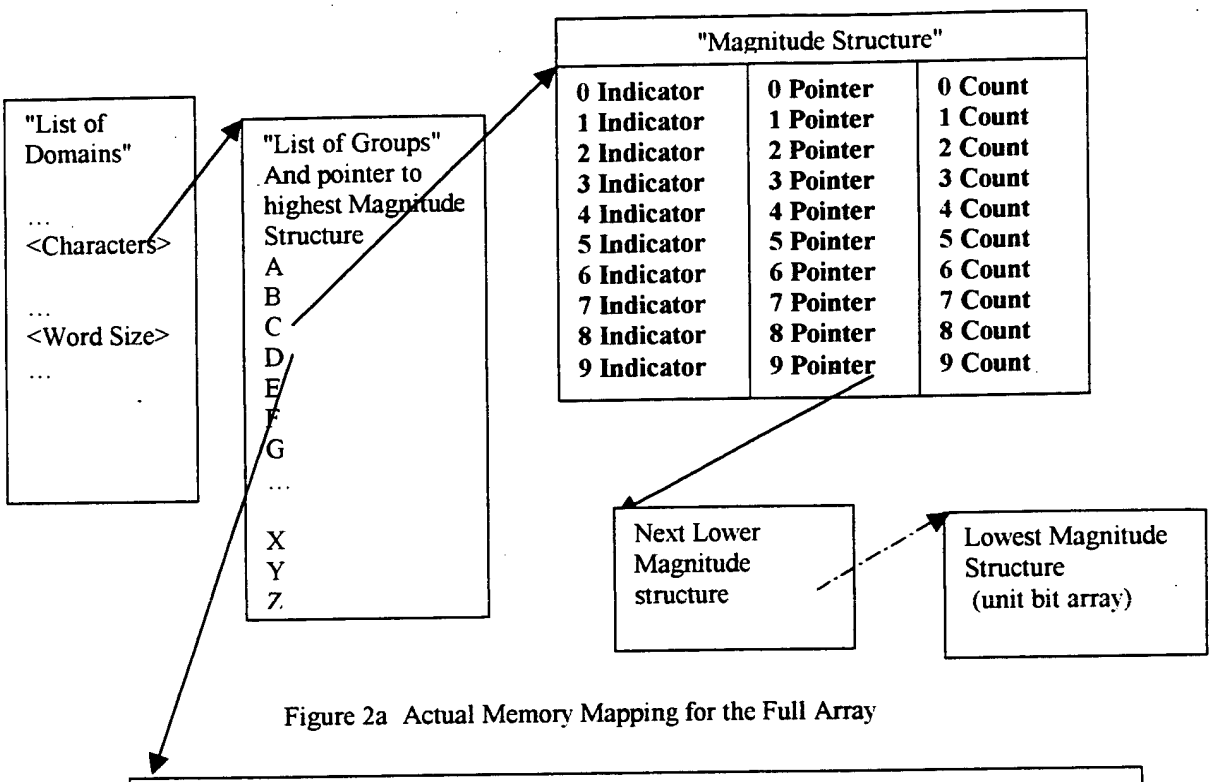


Figure 2a Actual Memory Mapping for the Full Array

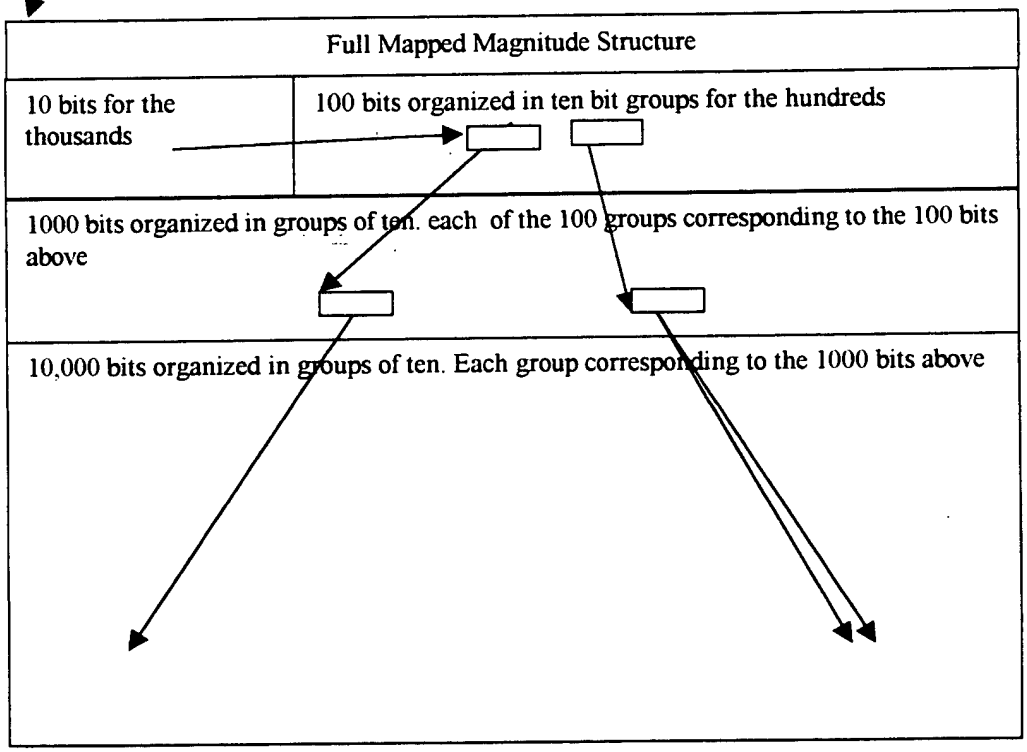
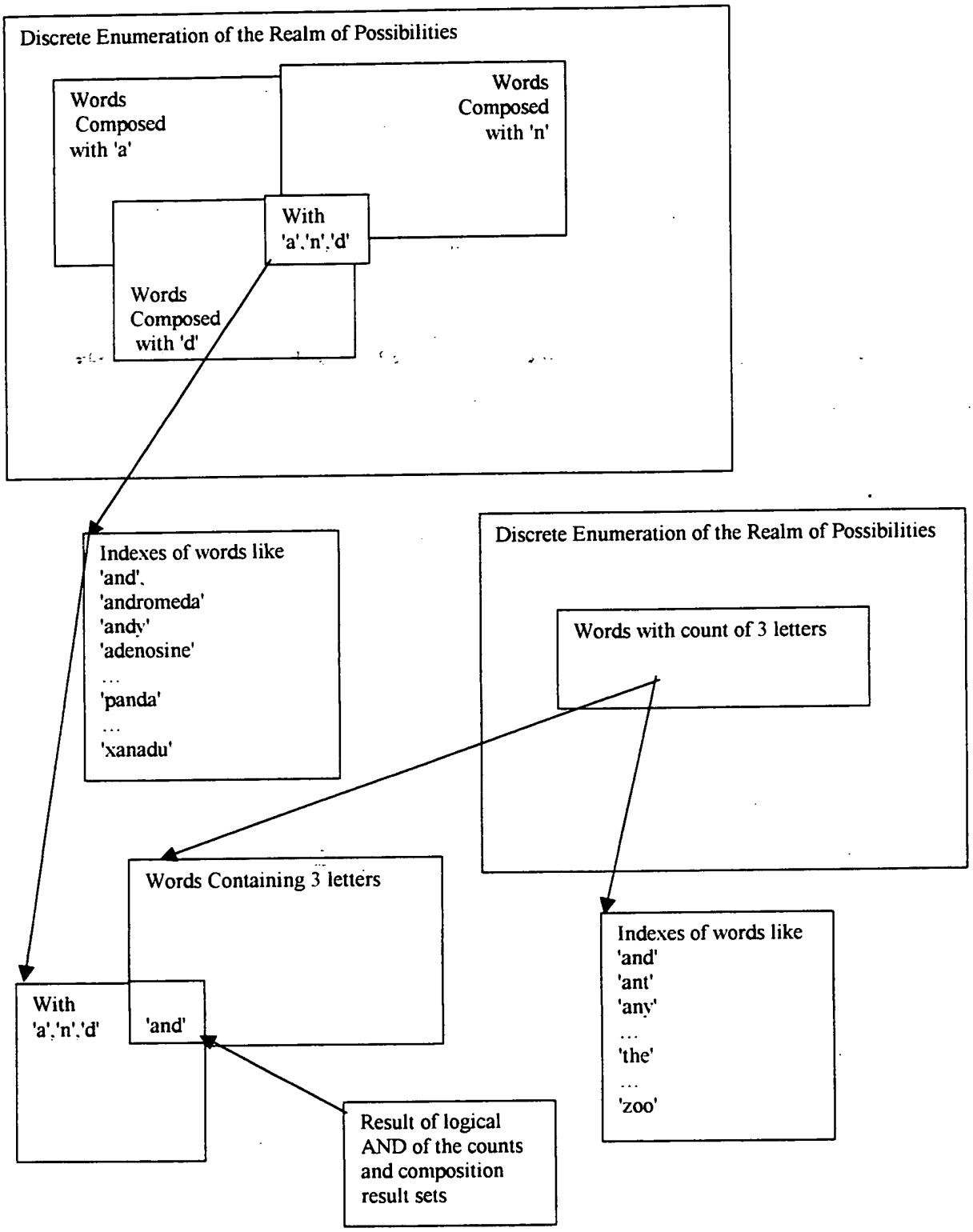


Figure 14


Discrete Enumeration of the Realm of Possibilities
Words Composed with 'a'
Words Composed with 'n'
With 'a','n','d'
Words Composed with 'd'
Indexes of words like
'and',
'andromeda'
'andy'
'adenosine'
...
'panda'
...
'xanadu'
Discrete Enumeration of the Realm of Possibilities
Words with count of 3 letters
Words Containing 3 letters
With 'a','n','d'
'and'
Indexes of words like
'and'
'ant'
'any'
...
'the'
...
'zoo'
Result of logical AND of the counts and composition result sets